UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/054,856	01/25/2002	Yuan Shao	01263.002303	8542

5514 7590 02/24/2005

FITZPATRICK CELLA HARPER & SCINTO
30 ROCKEFELLER PLAZA
NEW YORK, NY 10112

EXAMINER	
ALBERTALLI, BRIAN LOUIS	

ART UNIT	PAPER NUMBER
2655	

DATE MAILED: 02/24/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	10/054,856	SHAO, YUAN	
	Examiner	Art Unit	
	Brian L Albertalli	2655	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☐ Responsive to communication(s) filed on _____.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-25* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-25* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☒ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date *4/25/02*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

IDS

1. The Information Disclosure Statement file April 25, 2002 references an incorrect application number "09/462,808". The attached reference, "Signal Processing System" has been considered, but a proper IDS with the correct application number must be submitted in response to this action.

Furthermore, on page 1 of the Information Disclosure Statement file April 25, 2002, there is evidence that this IDS is a *second* IDS (see "Second Information Disclosure Statement" heading). However, there is no record of a first IDS for this case. If the applicant has additional references for the examiner to consider, they should be submitted with a proper IDS in response to this action.

Specification

2. The title of the invention is not descriptive. A new title is required that is clearly indicative of the invention to which the claims are directed.

The following title is suggested: --User Interface for Speech Model Generation and Testing--.

3. The disclosure is objected to because of the following informalities: there are no headings in the specification. The appropriate headings (Summary of the Invention, Brief Description of the Drawings, Detailed Description of the Preferred Embodiments, etc.) should be added at the appropriate points in the specification.

Appropriate correction is required.

Claim Rejections - 35 USC § 102

4. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

5. Claims 1, 10, and 25 are rejected under 35 U.S.C. 102(e) as being anticipated by Lewis et al. (U.S. Patent 6,826,306).

In regard to claim 1, Lewis et al. disclose an apparatus for generating and testing speech models (Fig. 1), said apparatus comprising:

a data collection unit (input unit 12 and memory 14) operable to collect and store utterance data indicative of the pronunciation of one or more words by one or more speakers (column 3, lines 35-40 and lines 46-49);

a speech model generation unit (training module 20) operable to generate speech models of words, utterances of which have been collected by said data collection unit (user develops user-dependent prototypes 22 and user-independent prototypes 24, column 3, lines 50-54 and lines 62-64); and

a testing unit (recognition engine 18, accuracy scores 26, and comparator 28) operable to test the accuracy of the matching of utterances collected by said data

collection unit to speech models generated by said speech model generation unit and to generate a visual display of the results of said testing by said testing unit (test data is used to test the decoding accuracy of the models, which is displayed on the graphical output unit 30, column 4, lines 10-17 and lines 30-34).

In regard to claim 10, Lewis et al. disclose a selector for selecting utterance data wherein said speech model generation unit is operable to generate speech models utilizing said utterances selected by said selector (a user selects a previous enrollment to generate the user-dependent prototypes used for recognition, column 7, lines 26-34).

In regard to claim 25, Lewis et al. disclose a method of generating speech models comprising the steps of:

providing a computer system operable to collect utterance data, to generate speech models utilising said collected utterance data and to test the accuracy of matching utterances to said generated speech models (Fig. 1);

collecting data indicative of the pronunciation of one or more words by one or more speakers utilising said apparatus (column 3, lines 35-40 and lines 46-49);

generating speech models utilizing said collected utterances (user develops user-dependent prototypes 22 and user-independent prototypes 24, column 3, lines 50-54 and lines 62-64);

determining whether said accuracy of said generated models is satisfactory by testing said models utilizing said apparatus (column 4, lines 10-17); and

outputting speech models determined to be satisfactory in said determination step (test data is used to test the decoding accuracy of the models, which is displayed on the graphical output unit 30, column 4, lines 10-17 and lines 30-34).

6. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

7. Claim 23 is rejected under 35 U.S.C. 102(b) as being anticipated by Gould et al. (U.S. Patent 5,850,627).

Gould et al. disclose a method of collecting utterance data comprising the steps of:

displaying a first user interface to enable user input of speaker identifiers and storing said speaker identifiers in a speaker database (Fig. 57, column 43, lines 27-30);

displaying a second user interface to enable user input of word identifiers and storing said word identifiers in a vocabulary database (Fig. 62, Add Word dialog box, column 46, lines 44-48);

displaying a series of prompts to prompt the utterance of words corresponding to word identifiers stored in said vocabulary database by speakers identified by speaker identifiers stored in said speaker database (the user interface of Fig. 61 prompts the user in element 1269 to pronounce words from that user's vocabulary and trains that user's model, column 45, lines 7-16); and

synchronising the collection of utterance data indicative of the pronunciation of words with said series of prompts (Fig. 61, after entering a user identifier through the interface of Fig. 57, two files are set up for each user that contain the identified user's vocabulary and models, column 43, lines 39-40; the user interface of Fig. 61 prompts the user in element 1269 to pronounce words from that user's vocabulary and trains that user's model, column 45, lines 7-16).

Claim Rejections - 35 USC § 103

8. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

9. Claims 2-4, 8, and 9 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lewis et al., as applied to claim 1, in view of Gould et al. (U.S. Patent 5,850,627).

In regard to claims 2 and 24, Lewis et al. disclose:

a vocabulary database operable to store word identifiers indicative of one or more words (a predetermined script is stored for the user to read while training the system, column 5, lines 6-14);

a speaker database operable to store speaker identifiers indicative of speakers from whom utterance data is to be collected (user profiles, column 5, lines 19-26).

Furthermore, Lewis et al. disclose a graphical user interface (output unit 30, column 4, lines 30-34) and that the user *reads* a series of prompts to train the models

(indicating that the prompts must necessarily be visually displayed by graphical user interface 30, column 5, lines 6-10).

Lewis et al. does not disclose the details of how the user interface is presented on graphical user interface 30.

Gould et al. disclose a system (Fig. 4) for training speech models wherein data is collected from a user with the aid of a graphical user interface. The system is operable:

to generate a first user interface to enable user input of speaker identifiers for storage in said speaker database (Fig. 57, column 43, lines 27-30);

to generate a second user interface to enable user input of word identifiers for storage in said vocabulary database (Fig. 62, Add Word dialog box, column 46, lines 44-48); and

to generate a third user interface operable to generate a series of prompts to prompt the utterance of words corresponding to word identifiers stored in said vocabulary database by speakers identified by speaker identifiers stored in said speaker database and to synchronize said series of prompts with the collection of utterance data indicative of pronunciation of words (Fig. 61, after entering a user identifier through the interface of Fig. 57, two files are set up for each user that contain the identified user's vocabulary and models, column 43, lines 39-40; the user interface of Fig. 61 prompts the user in element 1269 to pronounce words from that user's vocabulary and trains that user's model, column 45, lines 7-16).

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Lewis et al. to use the user interfaces disclosed by Gould et al. to

collect vocabulary and user information, and present prompts to the user to speak to collect, generate, and test the speech models, in order to provide an intuitive interface for a layperson to generate speech models.

In regard to claim 3, the user interface disclosed by Gould et al. (Fig. 61), as implemented in the combination of Lewis et al. and Gould et al. and applied to claim 2, above, comprises a generation of a series of visual instructions to speakers identified by speaker identifiers in said speaker database to pronounce words identified by word identifiers stored in said word database (each word in the vocabulary for each speaker is presented in element 1269, with the instructions "Please say" listed above, see Fig. 61 and column 45, lines 32-37).

In regard to claim 4, the user interface disclosed by Gould et al. (Fig. 61), as implemented in the combination of Lewis et al. and Gould et al. and applied to claim 2, above, comprises a prompt that only displays when a user is supposed to speak (column 45, lines 35-37). Not displaying word would be indicative of indicating to the user to stay quiet.

Neither Lewis et al. nor Gould et al. specifically disclose presenting a prompt to instruct the user to stay quiet preceding and succeeding instructions to pronounce a word.

Official notice is taken that it is notoriously well known and recognized in the art to indicate to a user to remain quiet preceding and succeeding instructions to

pronounce a word and to collect data throughout the display of those instructions, so that data indicative of the environmental noise where the user is speaking can be collected, and used to correct the user's model. Correcting the speaker model with information about the environmental noise greatly increases the recognition accuracy of the model.

It would have been obvious to one of ordinary skill in the art at the time of invention to further modify the combination of Lewis et al. and Gould et al. to specifically instruct a user to remain quiet, in order to collect information about the environmental noise, which could be used to correct the user's speaking model, which would increase the accuracy of the model.

In regard to claim 8, the user interface disclosed by Gould et al. (Fig. 61), as implemented in the combination of Lewis et al. and Gould et al. and applied to claim 2, above, comprises a selection unit operable to generate a user interface enabling user selection of speaker identifiers stored in said speaker database (Fig. 24, 418, user enters name and the user's corresponding vocabulary files are loaded, column 21, lines 20-29) and word identifiers stored in said vocabulary database (once the user has registered and is at the training interface of Fig. 61, the user selects words to train through word selection box 1261, column 45, lines 17-31) wherein said co-ordination unit is operable to generate a third user interface to generate a series of prompts to prompt the utterance of words corresponding to selected word identifiers by speakers corresponding to selected speaker identifiers selected utilizing said selection unit (only

the active training words selected through word selection box 1261 are presented in window 1269, column 45, lines 32-37).

In regard to claim 9, Lewis et al. disclose that when generating a user-independent model, data from a population of individuals is used using conventional training methods (column 3, line 62 to column 4, line 2). Conventional training methods generally require a user to repeat a word a number of times.

Official notice is taken that it is notoriously well known and recognized in the art that the generation of user-independent models requires multiple utterances of each word in a vocabulary from each of the plurality of individuals training the system.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Lewis et al. to determine the number of prompts shown to an individual training the speech models from the number of items of utterance data, so that an adequate amount of training data would be collected for each word model, thereby increasing the accuracy of the word model.

10. Claims 5-7 rejected under 35 U.S.C. 103(a) as being unpatentable over Lewis et al., as applied to claim 2, in view of Gould et al., and further in view of Nussbaum (U.S. Patent 5,867,816).

In regard to claims 5-7, nether neither Lewis et al. nor Gould et al. disclose means for presenting the collected utterance data back to the user.

Nussbaum discloses a system for developing and testing speech models. The system comprises means for:

displaying a waveform indicative of collected utterance data whilst said data is being collected (Fig. 1, element 56, speech data is displayed, see Fig. 9, Region 1 and column 6, lines 37-39); and

displaying a waveform and outputting audio data corresponding to the collected utterance data and permitting user deletion of stored utterance data output by said data collection unit (elements 56-58, the speech is displayed and audibly played back for the user to confirm that the input speech is adequate for training purposes, column 6, lines 37-47).

It would have been obvious to one of ordinary skill in the art at the time of invention to further modify the combination of Lewis et al. and Gould et al. to present the input utterance data both as a visual waveform and audibly, so that a user could confirm that the utterance data was acceptable, thereby ensuring that the speech models would not be modified with bad utterance data (such as noise or inarticulate speech).

11. Claims 11-14 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lewis et al., as applied to claim 10, in view of Ortega et al. (U.S. Patent 6,332,122).

In regard to claims 11-13, Lewis et al. disclose storing constraint data wherein the constraint data is gender data (a user profile is stored for each user wherein the user profile assigns the utterance data of each user to a male/female class, column 5, lines 19-32).

Lewis et al. do not disclose enabling a user to identify words and speakers fulfilling the requirements of the constraint data to generate speech models.

Ortega et al. disclose a system for training speech models in which a user identifies words (selects text, Fig. 5, 150) and speakers (speaker ID) to generate speech models utilizing said utterance data associated with said identified speakers and words (speech models are generated according to speech associated to the selected text and the identified user, column 6, lines 13-25).

Ortega et al. do not disclose identifying speakers which fulfill constraint data.

It would have been obvious to one of ordinary skill in the art at the time of invention to modify Lewis et al. to enable the user to identify words and speakers to utilize in generating speech models, so that poor speech data (from users with poor pronunciation, for example) would not be used in the training of the speech models.

Furthermore, official notice is taken that it is notoriously well known and recognized in the art that using speech with similar characteristics (constraint data) to train speech models produces more accurate speech models for speakers with those characteristics. For example, men's speech generally has a lower pitch and a different distribution of formants when compared to female speech, therefore, a speech recognition model trained exclusively with male utterance data will recognize male speech very well, but will not perform well at recognizing female speech. Similarly, a speech model trained with both male and female utterance data will tend to give equal, but mediocre, performance for both male and female speech.

It would have been obvious to one of ordinary skill in the art at the time of invention to further modify the combination of Lewis et al. and Ortega et al. to use the constraint data in the selection of words and users for training the speech models, so that user's with similar characteristics would be clustered together in the generation of the speech models, so the models would be more accurate for user's with that characteristic (male speech, which has different statistical properties than female speech, would not be allowed corrupt the models of the female speech and vice versa).

In regard to claim 14, neither Lewis et al. nor Ortega et al. disclose the constraint data is indicative of the number of repetitions of a word.

Official notice is taken that it is notoriously well known and recognized in the art that speech word models rapidly increase in accuracy as more repetitions of speech data is used to train the word models.

It would have been obvious to one of ordinary skill in the art at the time of invention to include constraint data indicative of the number of repetitions of a word in the selection of utterance data, so that speech models would only be created with utterance data with a sufficient number of repetitions required to ensure the accuracy of the speech model.

12. Claims 15-22 are rejected under 35 U.S.C. 103(a) as being unpatentable over Lewis et al.

In regard to claim 15, Lewis et al. disclose speech models are identified and used for testing (previous prototypes are stored and accessed at the user's discretion, column 7, lines 22-26). Furthermore, Lewis et al. disclose storing utterance data (column 3, lines 35-40 and lines 46-49).

Lewis et al. do not disclose allowing the user to select utterance data to test the speech models.

Official notice is taken that it is notoriously well known and recognized in the art to test speech models with utterance data from plurality of spoken utterances in order to ensure the speech models will be accurate over the slight changes in pronunciation for each instance of a user's utterance of a particular word.

It would have been obvious to one of ordinary skill in the art at the time of invention to allow the user to select utterance data to test the speech models in order to ensure the speech models will be accurate over the slight changes in pronunciation for each instance of a user's utterance of a particular word.

In regard to claim 16, Lewis et al. do not disclose the utterance data is indicative of the pronunciation of different words by different speakers.

Official notice is taken that it is notoriously well known and recognized in the art to test speech models with utterance data from a plurality of different speakers in order to ensure the accuracy of the model over the different pronunciation styles of the plurality of users. This ensures one model can accurately recognize a plurality of speakers.

It would have been obvious to one of ordinary skill in the art at the time of invention to allow the user to select utterance data from a plurality of different speakers to test the speech models in order to ensure the speech models will be accurate for a plurality of different users.

In regard to claim 16, Lewis et al. disclose testing models with utterance data collected from speakers used to generate the speech models (the accuracy of a user-dependent model is tested with data from that user, column 6, lines 1-4).

In regard to claim 17, Lewis et al. disclose testing models with utterance data collected from speakers not used to generate the speech modes (the accuracy of a user-independent model trained with a plurality of speakers is tested with data from a single user, column 5, lines 32-36).

In regard to claims 19-22, Lewis et al. disclose the disclosed system is implemented as software (column 4, lines 35-37), but does not explicitly disclose the storage medium storing the instructions thereon.

Official notice is taken that it is notoriously well known and recognized in the art to store software on a magnetic, optical, or magneto optical disk, or storing it as an electrical signal in a communications network.

It would have been obvious to one of ordinary skill in the art at the time of invention to store the software disclosed by Lewis et al., on a magnetic, optical, or

magneto optical disk, or store it as an electrical signal in a communications network, since these are all high density and long lasting means for storing computer implementable instructions.

Conclusion

13. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Kanevsky et al. (U.S. Patent 6,665,644) disclose a method for identifying speech data according to constraints, such as gender, age, accent, and dialect, and storing these results in an ordered speech utterance database. Nussbaum (U.S. Patent 5,867,816) disclose an operator interface for developing neural networks for the recognition of phonemes. Fado et al. (U.S. Patent 5,943,649) disclose a user interface that instructs a user to stay silent. Roberts (U.S. Patent 5,765,132) disclose a method for adding new words to a speech model database. Sneh (U.S. Patent 6,266,635) discloses a user interface for a user to enter a word, then enter speech to create a speech model for that word. Fado et al. (U.S. Patent 6,342,903) disclose user interface for collecting and storing speech data from a plurality of input devices for training data. Ortega et al. (U.S. Patent 6,675,142) disclose a method of testing the accuracy of transcribed speech. Aaron et al. (U.S. Patent 6,728,680) disclose the display of a waveform of speech helps a user recognizer whether the speech sample was acceptable.

14. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Brian L Albertalli whose telephone number is (703) 305-

1817 until March 24, 2005. After March 24, 2005, the examiner can be reached at (571) 272-7616. The examiner can normally be reached on Mon - Fri, 8:00 AM - 5:30 PM, every second Fri off.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Talivaldis Smits can be reached on (703) 305-3011. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

BLA 2/14/05

DAVID L. OMETZ
PRIMARY EXAMINER